SIMPLIFIED METAPHORICAL ANESTHESIA INTERFACES

Bachelorproject, 2010

Nino van Hooff, s1685848, a1685848@ai.rug.nl
Internal supervisor: Fokie Cnossen
Supervisor UMCG: Bert Ballast

ABSRTRACT: Objective: To show that anesthesiologists might be able to diagnose complications both faster and with higher accuracy when using a graphical histogram display instead of a classic cardiovascular display. **Method:** Our histogram displays shows the same 6 physiological variables as the cardiovascular monitor using colored bars rather than waveforms and numerical values. In a repeated measures design, 15 anesthesiologists were asked to make a diagnosis based on snapshots of both monitors. **Results:** Participants did not perform significantly better when using our histogram display. A strong effect was found however, of incidence on both accuracy and response time. **Discussion:** Although shorter response times were observed for the new display, the statistical power was not sufficient to confirm that the use of this alternative display caused the lower response times. An experiment with more subjects might still be able to confirm our hypothesis.

1. Introduction

In 1990, it was estimated that at least 2,000 preventable anesthetic mishaps occurred in the United States each year (Weinger & Englund, 1990). Ten years later, important steps had been taken towards increased patient safety, but there was still a lot of room for improvement. Vigilance in particular is still suboptimal. Easily detectable and correctable events are still missed, even when modern monitoring technologies are in place. (Gaba , 2000). Weinger & Englund (1990) assert that practicing anaesthesiology is a demanding job that imposes great cognitive load upon anaesthesiologists. A representation of the current situation has to be maintained, which requires integrating information obtained from the patient, various devices and other operating room personnel. Using this representation, the anesthesiologist has to constantly update hypotheses about the patient’s status. On top of that, attention must be paid to executing fine motor tasks such as intubation and laryngoscopy.

The field of Human Factors has applied itself to understanding human capability. The aim is to use the acquired knowledge to improve the design of machines and equipment. This is usually done by placing buttons where users expect to find them, and making information that requires special attention more salient. Another method is designing equipment to make inappropriate use impossible: shaping hoses so they will only fit on machines with similarly shaped connectors, for example (Gaba, 2000).

1.1. Patient monitors

The aim of the current study, performed at the University Medical Center of Groningen (UMCG), is to decrease the cognitive load experienced by anesthesiologists by simplifying the current patient monitors. As a reference, we use the Philips MP-70 IntelliVue monitor (Figure 1), which is currently in use at the UMCG. It displays graphs for ECG, Plethysmogram, Capnogram and Ventilation pressure.

Numerical values are displayed for heart rate, blood pressure (systolic, diastolic and mean), saturation (SPO2), end tidal CO2, PEEP, PAW and BIS value.

Figure 1: Philips MP-70 IntelliVue monitor.

1.2. Graphical monitors

Gurushanthaiah, Weinger, & Englund (1995) found that a change in data values was both more rapidly and more accurately detected when using graphical displays rather than numerical displays. This effect was not observed in nonmedical volunteers, however. Both a histogram and a polygon display were used. The histogram display depicted seven physiologic variables in the form of scaled linear tapes where a bobbin indicated the value of each variable on the vertical scale. The polygon display represented each variable as one vertex of a single geometric figure. The shape of the polygon changed proportionally as the value of each variable changed. Thus, as a variable increased or decreased in value, the corresponding vertex of the polygon enlarged outward or contracted inward, respectively (Figure 2). It was suggested that for the numerical display, the elements might be processed serially, while the graphic displays were processed holistically.

Blike, Surgenor, & Whalen (1999) show similar results with a graphical display of their own design. Accuracy and response time both improved significantly on trials concerning anesthesia-related shock detection.

Cole & Stewart (1993) proposed the 'rectangle metaphor' to present respiratory data. The width of the rectangle corresponds to the ventilation rate and the height of the rectangle to the tidal volume. By comparing two rectangles generated at different times, a change in either variable becomes apparent when the shapes of the rectangles differ.

The Integrated Graphical Anesthesia Display (IGAD) by Michels, Gravenstein, & Westenskow (1997) was one of the early attempts at replacing numerical patient monitors with a graphical alternative. No less than 30 variables were displayed on a single screen. The rectangles metaphor by Cole e.a. was used to display all data. A steady-state frame was added around the rectangles, however. This rectangle has a fixed size which corresponds to target values for the variable(s). The filled rectangle overflows when values are higher than expected. Apart from the physiological variables present on the numeric monitor discussed in section 1.1, the concentration or dosage of several anesthesia

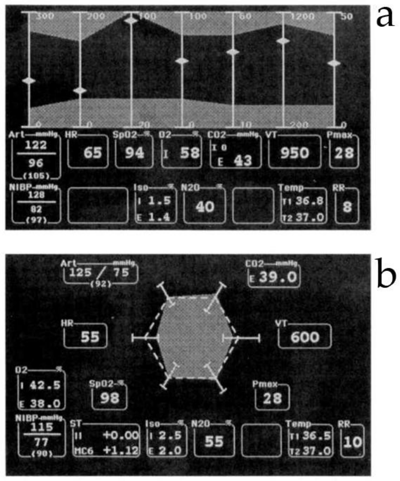

Figure 2: Histogram (a) and polygon (b) displays used by Gurushanthaiah, Weinger, & Englund (1995)

related drugs and fluids were displayed, as well as urine and blood loss. The IGAD display was compared with the BODY simulation software (Body Simulation, Advanced Simulation Corporation, San Clemente, CA). The IGAD group identified complications 2.8 to 3.1 minutes sooner than the BODY group did.

1.3. Metaphorical Anesthesia Interface

Combining aspects of these monitors, van Amsterdam (2010) created a monitor which he called Metaphorical Anesthesia Interface (MAI). From interviews with anesthetists, anesthesiology residents and nurse anesthetists, he learned that these practitioners did not consider fluids and drug dosage critical to diagnose anesthesia related complications. To be able to make a comparison with the current monitors, it was decided to include the same variables as those discussed in section 1.1. The variables O2 and CO2 as well as blood pressure and heart rate were stacked in one rectangle, because the interview showed that practitioners preferred this type of presentation. See Figure 4. Participants were asked to select 1 of 10 possible diagnoses based on snapshots of the monitors showing a complication. No significant difference in accuracy or response time between the MAI and classic monitor were observed however. Since van Amsterdam expected that subjects would benefit from more dynamic information, he added trend information to the MAI. Arrows showed the direction of a changing variable, as well as the speed of that change

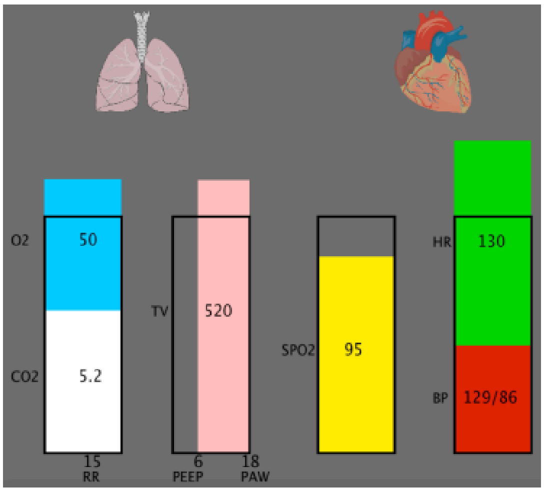

Figure 4: Metaphorical Anesthesia Interface (van Amsterdam, 2010).

(arrow thickness). No improvements in performance were observed however.

1.4. Current study

Since previous research has shown that graphical representations aid recognition accuracy and response time (Gurushanthaiah e.a., 1995; Michels e.a., 1997; Weinger & Englund, 1990) the results of van Amsterdam's study were unexpected. The aim of this study is to adapt the MAI monitor in order to improve the performance of its users.

When O2 and CO are stacked in one rectangle, as they are in the MAI monitor, they might fill up the rectangle when either is low and the other high. We suspect that in this case, one might be falsely led to believe that the patient is in a stable state. To prevent this, we represent all 6 main variables (O2, CO2, SPO2, TV, BP and HR) with one rectangle each.

Subjects participating in van Amsterdam's research commented that inspecting a monitor did not yield enough information to make a good diagnosis. We therefore consulted with an experienced anesthesiologist to make a list of complications which would be discernible by physiological data alone. This yielded a list of 9 complications. Details follow in the subsequent section.

We hypothesize that:

1. Complications will be recognized with greater accuracy when using our simplified MAI monitor, than with the classic monitor as described in section 1.1.
2. Complications will be recognized sooner when using our simplified MAI monitor, than the with classic monitor as described in section 1.1.

To test the hypotheses, we compared the performance of subjects in a repeated measures design using two types of monitors. The first monitor type was inspired by monitors which are currently being used by anesthesiologists. The other monitor is an alternative based on

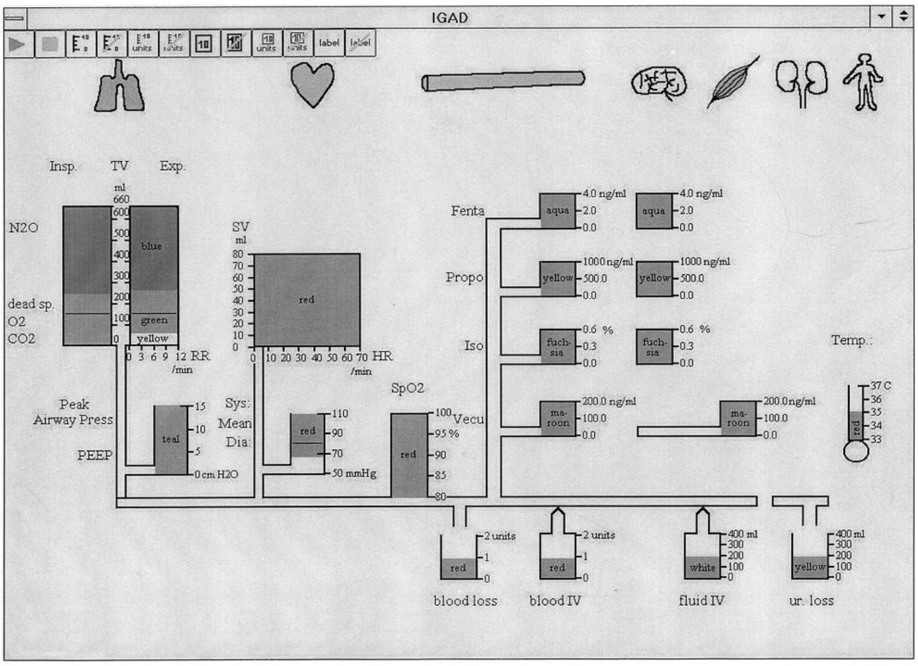

Figure 3: The IGAD display. 30 variables are displayed, using the rectangle metaphor by Cole e.a. The variables are organized according to function beginning on the left with respiratory, cardiovascular drug delivery, and fluid management.

principles and earlier work discussed in this section.

2. Method

2.1. Subjects

15 clinicians agreed to participate in the experiment. 14 were anesthesiologists while 1 was an anesthesiology resident. All subjects worked with both monitor types. Since the experiment was conducted during regular working hours, the schedule was adjusted to meet the participant's free time. Participants could be called back to the operation room at any time, but this did not occur. The experiment lasted 20 to 30 minutes, and subjects were not rewarded for their efforts. 4 out of 15 subjects had previous experience with the MAI monitor by van Amsterdam (2010).

2.2. Apparatus

A computer program was written in Java SE 6 to display monitor snapshots and to log answers and response times. The monitors show static information. Contrary to the experiment conducted with the IGAD by Michels e.a. (1997), no patient simulator was used. Therefore, participants were deprived of some additional information such as blood loss, cuff leakage and body temperature. The program was run on a 15.4" laptop equipped with a glossy screen. It was made sure that reflections in the screen did not cause any hindrance.

For the 'classic' monitor type, still images taken from a simulated anesthesia monitor were used. The monitor shows numerical data for all main variables, plus graphs for select variables. See Figure 6.

2.3. Metaphorical monitor

The simplified Metaphorical Anesthesia Interface (sMAI) presents information about the same variables. Each variable is represented by a colored bar. The fill color matches the color of the same variable in the classic monitor. The height of the bar is (linearly) proportional to the value of the variable. The stable state value of each variable is displayed as a black frame. Thus, if a value is too high, the bar will be higher than the black frame. If the patient is stable, all bars will be approximately level with their frames. The variables are not displayed in the same order as in the classic monitor. Instead, the six

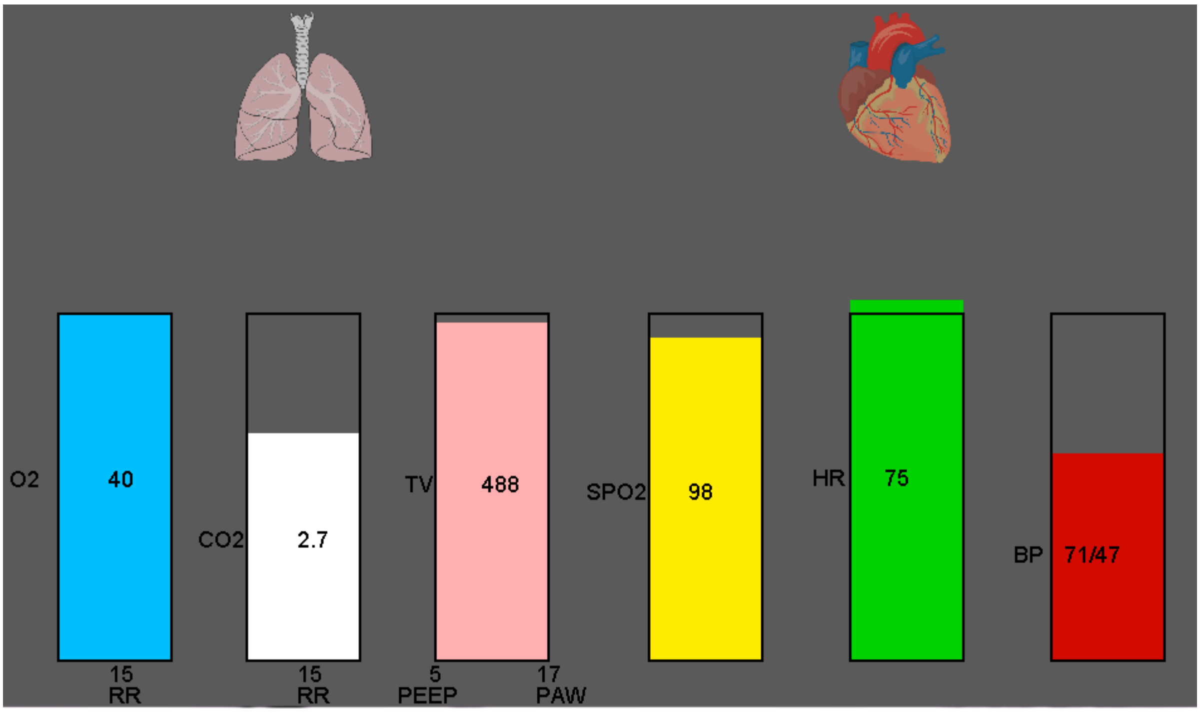

Figure 5: The simplified Metaphorical Anesthesia Interface (sMAI). Six physiological variable are grouped under pulmonary (left) and cardiovascular (right). The black outlines visualize the stable state. The colored rectangles may under- (value too low) or overflow(value too high) the outline. Numerical values are displayed in the center of each rectangle. Waveforms are not shown on the sMAI.

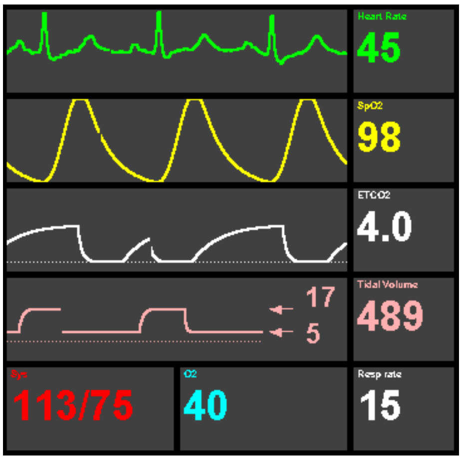

Figure 6: The classic monitor (simulation).

variables are subdivided in two categories: pulmonary and cardiovascular. This grouping has been made explicit by using an image of a set of lungs and a heart, respectively. The colored bars represent values relative to the stable state. The absolute values have been printed in the center of each bar. Note that the graphs displayed in the classic monitor are absent. See Figure 5.

2.4. Complications

In van Amsterdam's study, subjects diagnosed 42% of the complications incorrectly. (van Amsterdam, 2010). When selecting complications for our study, we wanted to make sure that all complications could be diagnosed correctly by observing the classic monitor only. A small pilot was conducted. All of the chosen complications (Table 1) were correctly identified by all 12 participating anesthesiologists. They were given only numerical data.

2.5. Procedure

2.5.1 Experiment

Since all tasks were completed with a computer mouse, subjects were first asked whether or not they were left-handed. The mouse was configured accordingly.
An experimental consisted of a snapshot of a complication which was represented by either the classic or the sMAI monitor. They were asked to select a diagnosis, after which they had to select one out of eight possible actions which they might want to execute first when they were confronted with the situation as depicted by the monitor. See Appendix A. The snapshots were generated from a hypothetical operation on a male patient, 42 years of age, 70kg (150 lbs.), and using pressure-constant ventilation.
Analyzing the first-action data goes beyond the scope of this thesis, and therefore they will not be evaluated.
The experiment started with a practice block, consisting of 3 trials. This block lets subjects get accustomed to using the program, the monitor, and the way answers are submitted. Only the heart rate varied, and the only possible diagnoses were bradycardia, stable and tachycardia. Since this was a practice round, a quick response was not required.

Then, to control for motor time, a second practice block was presented. The names of diagnoses were shown, and subjects searched for that name in the answer pane and clicked it. The names were shown twice, in random order. To generate random numbers, standard functionality of the Java SE 6 programming environment was used. Subjects were explicitly asked to answer as quickly as possible. The acquired motor times were used to distinguish between diagnosing time and motor time. See section 3.3.

The actual experiment was done in the final block. Subjects were confronted with 10 snapshots of the monitor. 9 corresponded to a complication, and 1 corresponded to the stable state. Although the participants were told that each complication could occur more than once, this was not the case. All 9 complications in Table 1 as well as the stable state were presented exactly once. For the stable-state values and the deviations from the stable state, see Table 2. The

Table 1: Participants were asked to select a diagnosis and then the action they would execute first to stabilize the patient.

Complication	Action
Anesthesia insufficient	Deepen anesthesia
Bradycardia	Atropine
Desaturation	100% O2
Hypertension	Deepen anesthesia
Hypoventilation	Incr. respiratory rate
Hypovolemia	Plasma substitutes
Air embolism	100% O2
Mal. Hyperthermia	Dantrolene
Stable	No action required
Tension pneumothorax	Thorax drain

Table 2: Complication features. "+" signifies an increased value and a "–" signifies a decreased value, compared to the stable state. "++" and "--" signify very large deviations from the steady state values, respectively. "# features" is the amount of deviating variables. Incidence data gives a measure of incidence. The number shown is the amount of reports out of 2000 (Webb e.a., 1993).

Complication	O2	CO2	TV	SP02	HR	BP	# features	Incidence
Stable	40%	4.1%	500 mL	100%	72 BPM	105 mmHg	0	-
Anesthesia insufficient					+	+	2	No data
Bradycardia					--		1	68
Desaturation				--			1	20
Hypertension						+	1	No data
Hypoventilation		+	-				2	47
Hypovolemia					+	-	2	No data
Air embolism		-				-	2	14
Mal. Hyperthermia		++		-	+		3	4
Tension pneumothorax				--		-	2	6

time it took subjects to select a diagnosis from the moment the snapshot was shown was recorded. Between each trial, a button was presented labeled "I am ready for the next one." This allowed participants to prepare themselves for the next trial.

2.5.2 Additional questions

12 out of 15 participants were willing to answer some additional questions. These questions concerned, among others, their preference for a certain monitor type and whether they had encountered all complications in the operation room before. We will discuss the results of this small questionnaire in the final section. The experiment lasted 20 minutes without, or 30 minutes including the optional questions.

2.6. Order of presentation

The advantage of a repeated measures design is that inter-subject variability is of lesser impact on the results. A disadvantage, however, is the possibility of a learning effect. To counter the recognition of complications while working with the second monitor, we made two datasets containing physiological data for the 9 complications plus stable state. The values in dataset 1 differ between 0 and 10 percent from the values in dataset 2. Since there are 2 datasets and two possible monitors to present first, 4 permutations are possible. We divided the permutations equally over the 15 subjects.

2.7. Measures

During the experiment, reaction times and motor times were logged. We reasoned that reaction times are a measure for recognition performance, although there might be between-subject variability in the motor times (eg. the time it took subjects to submit their answer after they had made their diagnosis). We used the lookup times obtained during block 2 of the experiment to correct the "raw" reaction times obtained during block 3. Since all diagnoses were presented twice, 2 lookup times were recorded for all diagnoses. The minimum of the two values was chosen as the lookup time to subtract from the raw reaction time. This resulted in the "corrected response time" which is referred to in the statistical analysis.
Answers to the diagnosis questions could be either correct or false. The number of correctly answered questions was assumed to be a measure for diagnosing accuracy.

3. Results

3.1. Internal validity

Since all subjects were tested with both monitors, and both monitors showed the same complications, a learning effect might occur. We would expect subjects to perform better on the last monitor they work with.
When using the classic monitor, the order in which the monitors were presented did not have a statistically significant effect on accuracy (mean

= 80% for both classic-first and sMAI-first; sd = 18% for classic first and 8% for sMAI first; t = .00 p < 1.00).

A paired- samples t-test was used to check for an effect of the used test set on response time. Although the variation of values was manually randomized, it might be easier for subjects to recognize a complication when working with either set rather than the other. Response times acquired with set 1 did not differ significantly from those acquired with set 2, however (t_{14} = 1.21; p< .25). No effect of sets was found on accuracy either (t_{14} = .48; p < .64). These results allow us to collapse all data regardless of the used data set or which monitor was used first.

3.2. Accuracy

Subjects performed equally well on average using the classic monitor (mean = 80%; sd = 14%) as they did using the sMAI monitor (mean = 78%; sd = 13%): the difference between monitors was not significant (t_{28} = 0.48; p < .69). Figure 8 shows that not all complication were diagnosed equally well. Table 3 shows which mistakes were made on the diagnosing task. The most diagnosing mistakes were made on air embolism, malignant hyperthermia and pneumothorax. Most diagnosing mistakes were

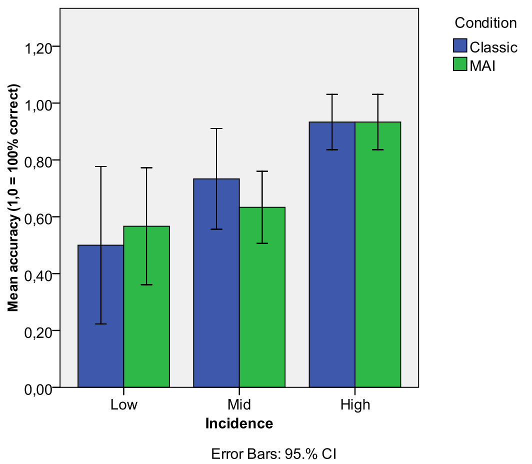

Figure 7: Accuracy vs. incidence. Incidence data from Webb e.a. (1993). A strong effect of incidence on accuracy was found.

unidirectional. Anesthesia insufficient and hypertension were interchanged by 4 subjects however.

We divided the complications into low- medium- and high frequency complications based on the incidence data from the Australian Incidence monitoring study (Webb e.a., 1993). For 6 of our complications, frequency data was available. Table 4 shows how the grouping was performed.

A repeated measures GLM (Huynh-Feldt

Table 3: Chosen diagnosis in case of incorrect answer.

Correct diagnosis \ Given diagnosis	Anesthesia insufficient	Bradycardia	Desaturation	Hypertension	Hypoventilation	Hypovolemia	Air embolism	Malignant Hyperthermia	Stable (no complication)	Tension pneumothorax	**Total (out of 30)**
Anesthesia insufficient		0	0	4	0	0	0	0	0	0	**4**
Bradycardia	0		0	0	0	0	0	0	1	0	**1**
Desaturation	0	0		0	1	0	0	0	0	0	**1**
Hypertension	4	0	0		0	0	0	0	0	0	**4**
Hypoventilation	0	0	0	0		0	0	1	0	2	**3**
Hypovolemia	0	0	0	0	0		0	2	0	1	**3**
Air embolism	0	0	0	1	2	10		0	0	5	**18**
Malignant Hyperthermia	2	0	1	0	7	1	0		0	1	**12**
Stable (no complication)	0	0	1	0	0	0	0	0		0	**1**
Tension pneumothorax	0	0	6	0	7	1	2	0	0		**16**
Total (out of 30)	**6**	**0**	**8**	**5**	**17**	**12**	**2**	**3**	**1**	**9**	

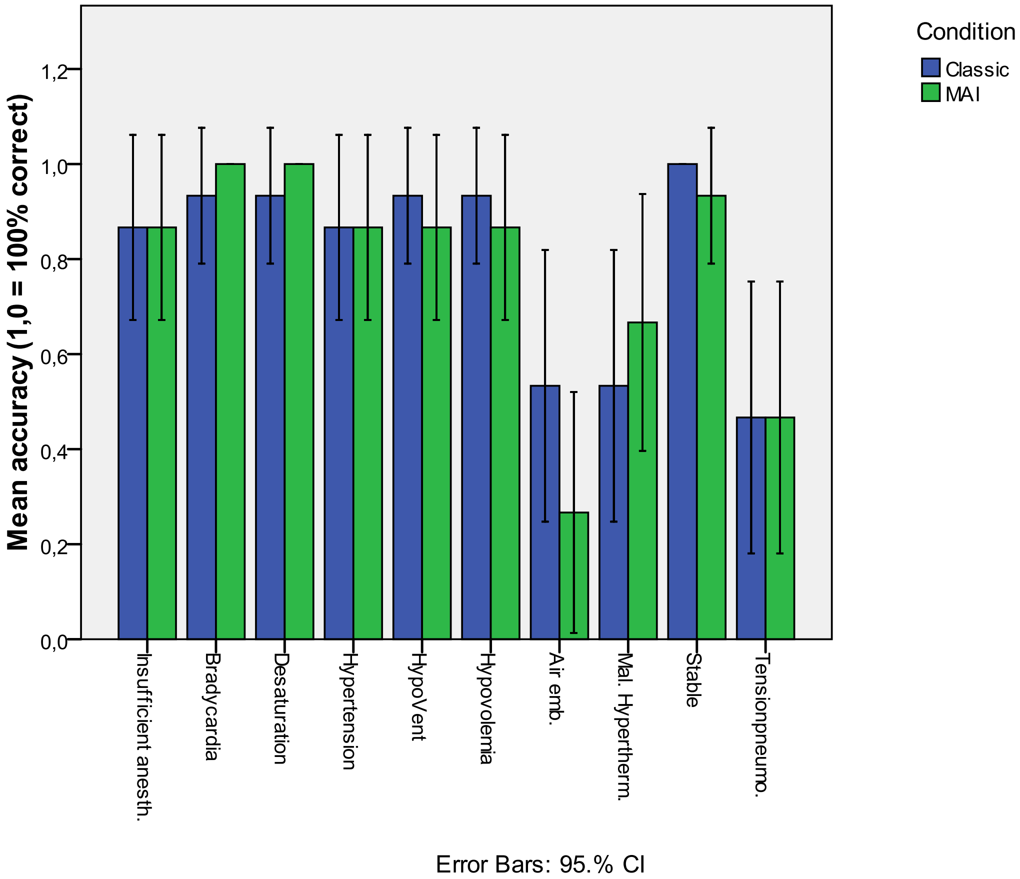

Figure 8: Accuracy contrasted between monitors. Labels indicate the correct diagnosis.

corrections) was used to test the effect of complication incidence on accuracy. A strong main effect was found (Type III sum of sq. = 2.45; df. = 1.8; F = 8.83; p < .01). No interaction effect on accuracy was found between incidence and monitor type (Type III sum of sq. = .11; df. = 2; F= 1.06; p< .36). Descriptives are summarized in Figure 7.

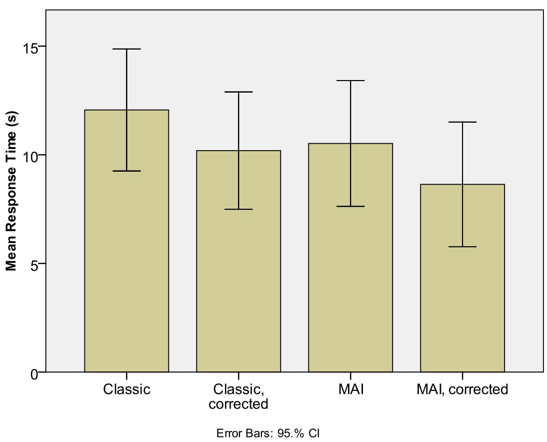

Figure 9: Measured response times with and without motor action correction.

3.3. Motor correction

A paired samples t-test shows that the corrected response times are positively correlated with the 'raw' response times (classic monitor: r=.999; p <.01, sMAI: r=.997; p <.01). The time it took a subject to select an answer (mean difference between corrected and uncorrected times) was 1871 ms (sd = 325 ms.) for the classic monitor and 1882 ms. (sd = 390 ms.) for the sMAI monitor. The difference between corrected and uncorrected times is not statistically significant (t_{14}=22.3 for classic and 18.7 for sMAI, p<.001 for both.). Therefore, we will only refer to uncorrected response times in the remainder of this thesis. Figure 9 shows the mean difference between corrected and uncorrected times.

3.4. Response times

We conducted a paired-samples t-test to assert whether monitor type has an effect on response time. We averaged the response times for all complications and clustered them according to

Table 4: Complications grouped according to incidence. The number between parentheses indicates the number of reports out of 2000 reports collected by Webb e.a. (1993).

Group	Members
Low incidence	Malignant hyperthermia (4), Tension pneumothorax (6)
Medium	Air embolism (14), Desaturation (20)
High incidence	Hypoventilation (47), Bradycardia (68)

monitor type. Response times for the sMAI monitor where shorter (mean = 10.5s; sd = 5.2s) than the response times for the classic monitor (mean = 12.0s; sd = 5.1s). The effect was not statistically significant ($t = 1.64$, $p < .12$). See Figure 12.

Figure 10 shows that response times were not consistently shorter for the sMAI monitor. For 4 out of 10 complications the response time for the sMAI was higher on average than for the classic monitor.

A repeated measures GLM with Huynh-Feldt corrections shows a strong main effect of incidence on response times. More frequently

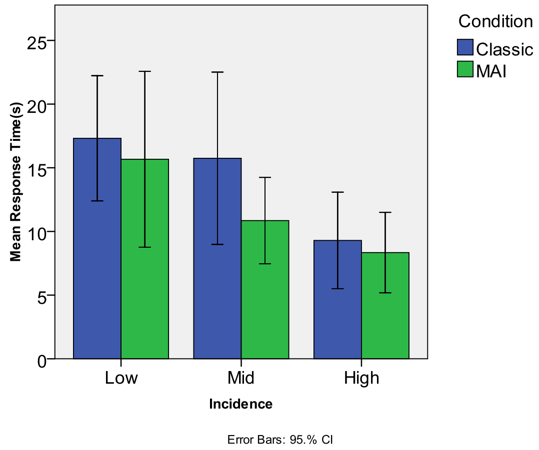

Figure 11: response times (seconds) vs. incidence, according to Webb e.a. (1993). Common complications were recognized faster.

occurring complications predicted higher response times (Type III sum of sq. = 8.92E8, df. = 1.7, $F = 7.25$, $p < .01$). No interaction effect was found between incidence and monitor type on response time (Type III sum of sq. = 6.63E8, df. = 1.8; $F = .7$; $p < .49$). See Figure 11.

4. Discussion

In this study we compared the performance of anesthesiologists using the simplified MAI monitor with their performance using the monitor which is in use today (the 'classic

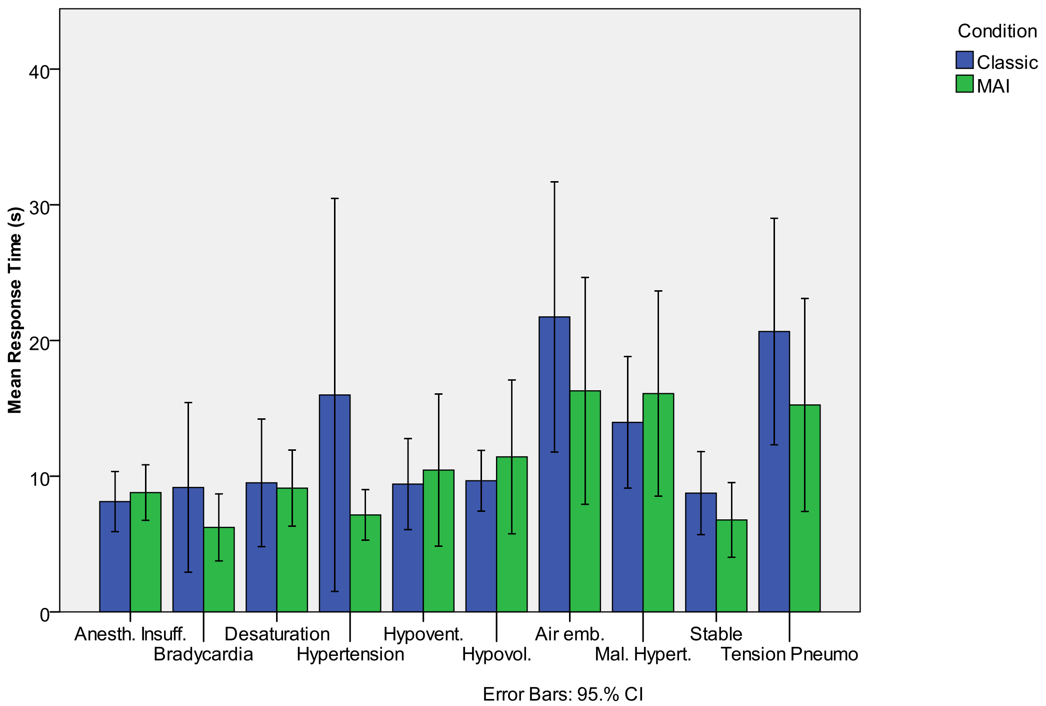

Figure 10: Uncorrected response times per complication. Labels indicate the correct diagnosis.

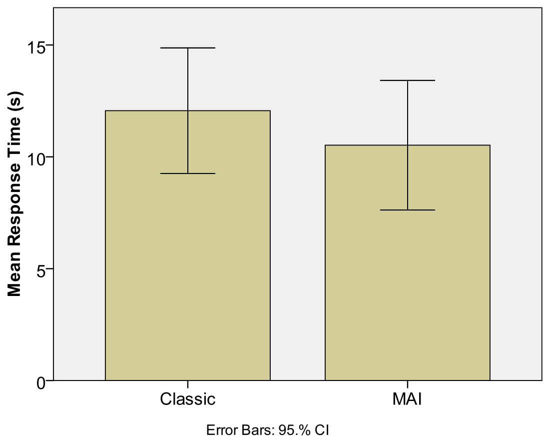

Figure 12: Response times in seconds contrasted between monitor types. The difference between sMAI and classic monitor types is not significant.

monitor'). It was assumed that response time was an appropriate measure for recognition performance and accuracy an appropriate measure for diagnosing performance. We tested the following hypotheses:

1. Complications will be recognized with greater accuracy when using our simplified MAI monitor, than with the classic monitor as described in section 1.1.
2. Complications will be recognized sooner when using our simplified MAI monitor, than with the classic monitor as described in section 1.1.

In the following subsections, these hypotheses will be evaluated.

4.1. Accuracy

No effect of monitor type on accuracy was found. It must be noted however, that most subjects had multiple years of experience with the classic monitor while they had only one hour of experience with the sMAI monitor. This might indicate that with more training, anesthesiologists might be able to achieve higher accuracy. We have not tested this however, so for the time being, we cannot confirm our first hypothesis.

A strong effect was found of incidence rate on accuracy (Figure 7). When thinking of incidence and medical reasoning, the well known availability heuristic comes to mind. Tversky & Kahneman (1974) argued that people will consider an event more likely when they can more easily recall examples of such an event from their experience. Applying this theory to the current study, we would expect anesthesiologists' diagnoses to be biased towards the more frequently occurring complications. Looking at our data, we see that out of 12 incorrect diagnoses given for tension pneumothorax, hypoventilation was diagnosed 7 times. Similarly, malignant hyperthermia was misdiagnosed as hypoventilation 7 times as well, out of 16 misdiagnoses (Table 3). It is noteworthy that both tension pneumothorax and malignant hyperthermia are observed much less frequently than hypoventilation (Table 4). This particular example supports the hypothesis that using this heuristic is responsible for a reduced diagnosing accuracy. Our incidence data is incomplete however, so we cannot rule out that counterexamples could be given.

Another possible explanation for low accuracy on some complications is that participants were not able to distinguish between complications. It seems quite likely to be the case for the misdiagnosis of hypertension as insufficient anesthesia, which is bidirectional (Table 3). Both complications are characterized by high blood pressure. An increased heart rate is only indicative of insufficient anesthesia, however. All other misdiagnoses are unidirectional. *Hypertension-insufficient anesthesia* and *Desaturation-Tension pneumothorax* are the pairs of complications which are the most similar, with the same features save for the omission of a single feature. Tension pneumothorax was misdiagnosed as desaturation 6 times. It is advisable to avoid this kind of near-ambiguities in future experimental designs. Since participants were obviously exposed to the same complications when using the sMAI as when using the classic monitor, this does not skew the results in favor of either monitor, however.

4.2. Response times

No effect of monitor type on response times was found. This means that we cannot confirm hypothesis 2 either. Where the accuracy between monitor types was almost identical for the classic monitor and the sMAI, the response times obtained with the sMAI monitor were 12.5% shorter than those obtained with the classic monitor. The confidence interval was too great for statistical significance, however (Figure 12).

With more participants, an effect might still be found, due to increased statistical power.
As with accuracy, an effect was found of incidence rate on response time, where lower incidence predicted higher response times.
When looking at the response times obtained from the stable state, we observe shorter response times for the sMAI monitor. This was expected since we suspect that the sMAI might be mentally processed as a whole, while the variables of classic monitor might be processed serially (Gurushanthaiah e.a., 1995). In case of the stable state, all sMAI bars are roughly lined up with their frames, and it can be asserted that all is well with a single glance at the monitor. Note that a larger sample size might be required to make the shorter response times for the sMAI monitor statistically significant.

4.3. van Amsterdam's research

The current research was conducted in close collaboration with van Amsterdam, and might be considered a follow-up study of van Amsterdam (2010). The obtained results of both studies are similar to one another. Diagnosing accuracy and response times were not significantly improved by either sMAI or MAI when compared to the classic monitor. When comparing studies however, we obtained a much higher diagnosing accuracy for both classic and metaphorical monitors. Since the classic monitor was unchanged, we suspect that the little pilot study we conducted to pick complications which are easily discernable from one another is responsible for this increased performance. With the adjustments we made to the monitor and the experimental setup we have also achieved an improvement in response time, from 3.6% (MAI) to 12.5% (sMAI). Again, these results are not statistically significant. The improved response times might be explained by a) more suitable complications and/or b) an easier to interpret monitor.

4.4. Further research

We have demonstrated a strong effect of incidence rate on both diagnosing accuracy and recognition time of anesthesiology-related complications. We have postulated the availability heuristic as a possible cause for this phenomenon, but it is outside the scope of the current research to ascertain this claim. A literature study on anesthesia-related incidence should show interesting results when combined with an experiment such as our own.
When comparing the sMAI to the IGAD display by Michels e.a. (1997), it is surprising to see that the different results were obtained. The sMAI was derived from the IGAD by removing all information which does not appear on classic monitors. It is possible that participants needed this information to make a diagnosis, but interviews with participants indicate otherwise.

The majority of participants of both our own and van Amsterdam's experiment complained about the static nature of the task. They thought that the developmental history of a complication would give them valuable information to make a better classification. We noticed that anesthesiologists cannot always rely on this kind of information since they regularly leave the operating room after the patient has been sedated. When a complication occurs, they are called back and can only rely on the information provided by nurses and other personnel. But, it is indeed a key difference in experimental design that the IGAD display was driven by a running simulation while our display presented a snapshot of a moment in time. While our experiment was a task of categorization, Michels' experiment was more a task of detecting changes in variables as time progresses. Gurushanthaiah e.a. (1995) showed that graphical representations aided both the accuracy and response time of perceived change. We thus strongly recommend a dynamic setting for further experiments of this type.

5. Acknowledgements

I would like to thank a few people for helping me conduct this research.
Firstly, I am indebted to my supervisor at the *Rijksuniversiteit Groningen* Fokie Cnossen, who offered me this bachelorproject and put me in contact with the University Medical Center of Groningen (UMCG). She helped me with the experimental design and the proper way to format this thesis. I am especially grateful for her patient explanation of the many statistical methods.

Secondly, I would like to thank Bert Ballast, external supervisor and anesthesiologist at the UMCG, for conducting a small pilot to select suitable complications for our experiment. I would also like to thank him for recruiting his colleagues as subjects and lending me his insights in the outcome of the experiment.
Finally I would like to thank Kai van Amsterdam, who generously shared the source code for the MAI monitor, on which the sMAI is based. It saved me a lot of work.

References

van Amsterdam, K. (2010). *Integration of Visual Methaphors in an Anesthesia Monitor*. Master thesis, Groningen. Obtained from: http://scripties.fwn.eldoc.ub.rug.nl/FILES/scripties/Kunstmatigeintellige/MasterHuman-machinec/2010/Amsterdam.K.van./AI-MHMC-2010-K.van.Amsterdam.pdf

Blike, G. T., Surgenor, S. D., & Whalen, K. (1999). A graphical object display improves anesthesiologists' performance on a simulated diagnostic task. *Journal of Clinical Monitoring and Computing*, *15*(1), 37-44.

Cole, W. G., & Stewart, J. G. (1993). Metaphor graphics to support integrated decision making with respiratory data. *International Journal of Clinical Monitoring and Computing*, *10*(2), 91-100.
Gaba, D. M. (2000). Anaesthesiology as a model for patient safety in health care. *BMJ*, *320*(7237), 785-788. doi:10.1136/bmj.320.7237.785

Gaba, D. M. (2000). Anaesthesiology as a model for patient safety in health care. BMJ, 320(7237), 785-788. doi:10.1136/bmj.320.7237.785

Gurushanthaiah, K., Weinger, M. B., & Englund, C. E. (1995). Visual Display Format Affects the Ability of Anesthesiologists to Detect Acute Physiologic Changes. *Anesthesiology*, *83*(6), 1184-1193. doi:10.1097/00000542-199512000-00009

Michels, P., Gravenstein, D., & Westenskow, D. R. (1997). An integrated graphic data display improves detection and identification of critical events during anesthesia. *Journal of Clinical Monitoring*, *13*(4), 249-259.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124-1131.
Weinger, M., & Englund, Carl. (1990). Ergonomic and Human Factors Affecting Anesthetic Vigilance. *Anesthesiology*.

Weinger, M.B. & Englund, C.E. (1990). Ergonomic and Human Factors Affecting Anesthetic Vigilance and Monitoring Performance in the Operating Room Environment. Anesthesiology, 73, 995-1021.

Webb., R.K., Van der Walt., J.H., Runciman, W.B., Williamson, J.A., Cockings, J., Russel, W.J. & Helps, S. (1993). Which Monitor? An Analysis of 2000 Incidents Reports. Anaesthesia and Intensive Care. 21, 529-542.

Appendix

A.

Figure A.1: The sMAI monitor and answer pane as presented to the subject. Subjects are asked to select a diagnosis in the answer pane (in Dutch). The correct answer for this example is air embolism ('luchtembolie').

Figure A.2: After a diagnosis is selected, the monitor disappears and the options in the answer pane are replaced by a number of anesthesiology-related actions. The task is to pick the one which will remedy the previously shown situation. The correct answer for air embolism is '100% O2'.

Figure A.3: Finally, a button is presented labeled 'ready for the next one'. This allows participants to prepare themselves for the next trial. After the button is clicked, a new monitor snapshot is presented, as shown in Figure A.1.